TP248 P77.P763

STIC-ILL

From: Ly, Cheyne
Sent: Saturday, April 12, 2003 12:59 PM
To: STIC-ILL
Subject: ILL_Order

1. J Aqvist, C Medina, and JE Samuelsson

A new method for predicting binding affinity in computer-aided drug design
Protein Eng. 1994 7: 385-391.

1. MA Saqi and JM Goodfellow

Free energy changes associated with amino acid substitution in proteins
Protein Eng. 1990 3: 419-423.

C. Dune Ly
Patent Examiner
Art Unit 1631
Room 12D-04
Mailbox 12D-01
Phone: 703 308 3880
Fax: 703 749 9331

Protein Engineering vol.3 no.5 pp.419–423, 1990

Free energy changes associated with amino acid substitution in proteins

Mansoor A.S.Saqi and Julia M.Goodfellow

Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

The estimation of free energy differences from computer simulation of macromolecular systems is important for rational strategies for drug design and for protein engineering. As an example of one mutation, we have studied the free energy change resulting from the conversion of a polar group (OH) to an apolar group (CH_3) in aqueous solution. We have estimated the effect of various local environments on the magnitude of the free energy difference and find that significant environmental effects are found. We have also studied the reliability of the results in detail.

Key words: free energy simulations/Monte Carlo/perturbation method

Introduction

Recently much effort and computer time has been spent on free energy perturbation (FEP) simulations, both for small molecules and for larger biological systems (Jorgensen, 1989; Beveridge and Di Capua, 1989). Such calculations have been described as the 'Holy Grail of theoretical chemistry' (Pearlman and Kollman, 1989a). Whilst the statistical mechanics of the FEP method are straightforward, the main problem, inherent in all simulations, but particularly in FEP calculations, remains: that of adequately sampling the phase space.

Despite their limitations, free energy calculations are exciting and have attracted much attention because of their potential. Experiments in protein engineering, often involving a small structural change, can lead to dramatic differences in properties such as enzyme activity and substrate binding (Thomas *et al.*, 1985). The change in free energy is the key index for stability and so estimation of this, using simulation techniques, is very important for the development of rational strategies for protein engineering and drug design.

The FEP method has been used to study the differences of Cl^- and Br^- solvation (Lybrand *et al.*, 1985) and the much larger change, that of CH_3OH to CH_3CH_3 (Jorgensen and Ravimohan, 1985). This interconversion of methanol to ethane in aqueous solution has now become one of the standard tests of free energy simulation methods. The calculation of free energies for mutations involving large perturbations introduces a number of problems. The mutation cannot be carried out in one step and must be broken down into several steps, and the free energies accumulated to give the total free energy change. The exact method and the length of the simulation at each interval can affect the final answer.

Singh *et al.* (1987) have studied, using molecular dynamics, a large structural change, viz. the interconversion of phenylalanine to alanine in aqueous solution. Although there are no experimental data for the free energies of solvation of alanine and phenylalanine, comparison with data on representative model systems for the sidechains (Wolfenden *et al.*, 1981; Ben-Naim and Marcus, 1984) shows good agreement.

Bash *et al.* (1987a) have performed a series of free energy calculations involving many mutations of amino acid side chains. The encouraging results of these simulations have prompted similar calculations on more specific biological systems, such as free energy of binding and of activation for catalysis of a tripeptide substrate by native subtilisin and a subtilisin mutant formed by changing Asn155 to Ala155 (Rao *et al.*, 1987). Although only small differences are found in the binding free energy, the corresponding catalytic free energy is substantial and in good agreement with experimental results. This method has also been used to study the relative free energy of binding, of the enzyme thermolysin with two inhibitors (Bash *et al.*, 1987b).

Brooks (1989) has calculated relative free energies of binding of various drugs to dihydrofoliate reductase. The free energy difference when TET (triethyl substituted cogener) exchanges with trimethoprim is estimated to be $+1.9 \pm 0.7$ kcal/mol, which can be compared with an experimental value of -1.7 kcal/mol. This difference implies that errors in sampling and in the potential parameters can have a significant effect and may be important in estimating overall differences in free energy.

In addition to simulations on specific systems of biological interest, there is a need for more systematic studies on smaller systems to see if any general trends emerge. Assuming that accurate and precise estimates of free energy differences can be calculated, we wish to study the effects of local environment on the magnitude of the free energy differences. Initial FEP calculations were undertaken on the mutation of ethane to methanol in aqueous solution (Jorgensen and Ravimohan, 1985; Brooks, 1989). This involved the mutation of a methyl to an hydroxyl group. Does the same mutation lead to a similar free energy difference if we study valine mutating to threonine in solution or in a polypeptide chain in solution, and is a similar simulation protocol suitable?

In an attempt to answer these questions about the effect of the local environment, we have studied the change of a polar group (OH) to an apolar group (CH_3) for a methanol to ethane mutation and a threonine to valine mutation in an Ala-Thr-Ala tripeptide and on Ala-Lys-Thr-Lys-Ala pentapeptide, both in the α helical conformation. We have also looked in detail at the precision and accuracy of our estimates.

Computational procedures

The difficulty in estimating free energies from simulation calculations is well known (Mezei and Beveridge, 1986) and arises from the fact that the free energy is not an ensemble average. The free energy difference between two states i and j can, however, be expressed in terms of an ensemble average by

$$F_{ij} = kT \ln \langle e^{-\beta U_n} \rangle_i \qquad (1)$$

Table I. Methanol to ethane mutation

Coupling parameter		F_{ij}	F_{ji}
λ_i	λ_j		
0	0.125	–	−2.086 ± 0.133
0.125	0.25	2.212 ± 0.123	−1.858 ± 0.094
0.25	0.5	2.263 ± 0.213	−1.858 ± 0.097
0.5	0.75	0.835 ± 0.091	−0.926 ± 0.102
0.75	1.0	0.282 ± 0.092	–

The forward (F_{ij}) and reverse (F_{ji}) free energies in kcal/mol. Equilibration was for 400 000 Monte Carlo steps and data was collected in batches of 50 000 for the next 1.0×10^6 steps.

where i is an initial state, j is the final state, β is $1/kT$ and $U_{ij} = U_j - U_i$, where U_i and U_j are the configurational energies of states i and j respectively.

Whilst equation (1) is exact, a meaningful estimate of the ensemble average can only be made if the perturbing potential U_{ij} is small. The ensemble average is then given by;

$$\langle e^{-\beta U_{ij}} \rangle_i = \frac{\{e^{-\beta U_{ij}} e^{-\beta U_{id}} N_q}{\{e^{-\beta U_{id}} N_q} \quad (2)$$

The reason that the perturbing potential (U_{ij}) has to be small is due to the sampling being based on a Boltzmann distribution of the reference state i. If the perturbation is too big those configurations chosen [i.e. where exp $(-\beta U_i)$ is large] are precisely the ones where exp $(-\beta U_{ij})$ is small. Hence the configurations which contribute most to the integral in equation (2) may not be sampled during the course of a simulation.

The free energy change is, however, a path-independent quantity and so the total free energy change between an initial (0) and final (1) state can be obtained by summing the free energy changes over a number of intermediate non-physical states, described by a hybrid hamiltonian. The system is smoothly transformed from state 0 to state 1 with the aid of coupling parameter λ. The number of these sub-states (the number of intervals or 'windows') and the size of each window are important in obtaining realistic results because the free energy change does not vary linearly with λ.

The hybrid hamiltonian is described by

$$H(\lambda) = \lambda H_1 + (1 - \lambda) H_0 \quad (3)$$

and the geometry of the system is also varied in a similar fashion.

Our Monte Carlo simulations have been run on a CRAY X-MP 28 (at the University of London Computer Centre) using the program BOSS (W.L.Jorgensen, Purdue University, Indiana), which includes double-wide preferential sampling and a feathered cut-off between 8.0 and 8.5. Double-wide sampling means that free energy differences for $\lambda_i \rightarrow \lambda_{i+1}$ and $\lambda_i \rightarrow \lambda_{i-1}$ can be obtained from one simulation, since in both cases the sampling is based on the λ_i distribution.

In this study we use five windows, with λ = 0.0, 0.125, 0.25, 0.50, 0.75 and 1.0 for both the methanol/ethane mutation and the threonine/valine mutation. An important check on the precision of the simulation is to compare F_{ij} and F_{ji}. Ideally, $F_{ij} = -F_{ji}$. The discrepancy between the forward (F_{ij}) and backward (F_{ji}) simulation is known as the hysterisis and is due to inadequate sampling caused by too large a perturbing potential. Observation of the hysterisis effect for each window enables us to highlight the difficult windows, which should be sub-divided into smaller intervals.

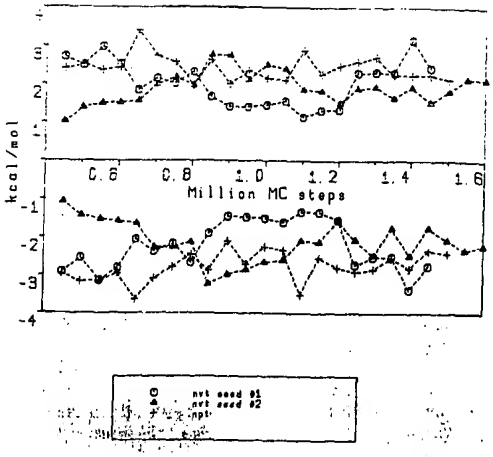

Fig. 1. Convergence profile for the Ala-Thr-Ala → Ala-Val-Ala mutation. Batch averages of free energy are plotted for two (N, V, T) simulations with different starting seeds and for an (N, P, T) simulation. The data are for the interval 0.125 → 0 (lower plot) and 0.125 → 0.25 (upper plot).

The system consists of the solute in a cubic box of side length 18.63. The solute atoms are described by the OPLS force field (Jorgensen and Tirado Rives, 1988), which employs a united atom approach where hydrogens attached to carbons are not represented explicitly. The number of water molecules is estimated using the partial molar volumes for the solute (Zamytnin, 1984) and they are modelled using the TIP4P model (Jorgensen, 1985).

In the first calculation, an N, V, T simulation of the interconversion of methanol to ethane is carried out. Standard geometries are used for ethane and methanol, and the molecules are kept fixed in these geometries during the simulation. The oxygen of methanol was converted into the CH_3 group of ethane and the hydrogen of methanol gradually disappears. The C–O bond length of 1.43 for methanol grow to the C–C bond length of 1.53 for ethane. 0.4×10^6 MC-steps of equilibration are used and data are collected in batches of 50 000 for the next 1×10^6 MC steps.

A similar protocol is used for the N, V, T simulation of the mutation of threonine to valine, within an Ala-Thr-Ala tripeptide in the α helical conformation with the X_1 torsion angle for the set at 288° (McGregor *et al.*, 1987). The OH of the threonine side-chain is mutated to a CH_3 group of the valine. There is no bond growth, and the geometry of the system remains fixed. For this system, the input to the BOSS is modified such that the coordinates, initial atom types and final atom types are handled directly to the program rather than the more usual z-matrix style input.

We repeated the calculation for the first two windows $\lambda = 0 \rightarrow \lambda = 0.125 \rightarrow \lambda = 0.25$ for the threonine to valine mutation in the tripeptide by changing the starting seed. This allows the system to take a different walk through phase space, and helps to give some indication of the adequacy of sampling. We also performed an NPT simulation for this window.

We then ran an N, P, T simulation of the mutation of threonine to valine but this time as part of the pentapeptide Ala-Lys-Thr-Lys-Ala again in the helical conformation. The torsion angle, X_1 for lysine was taken as 287° (McGregor *et al.*, 1987) and

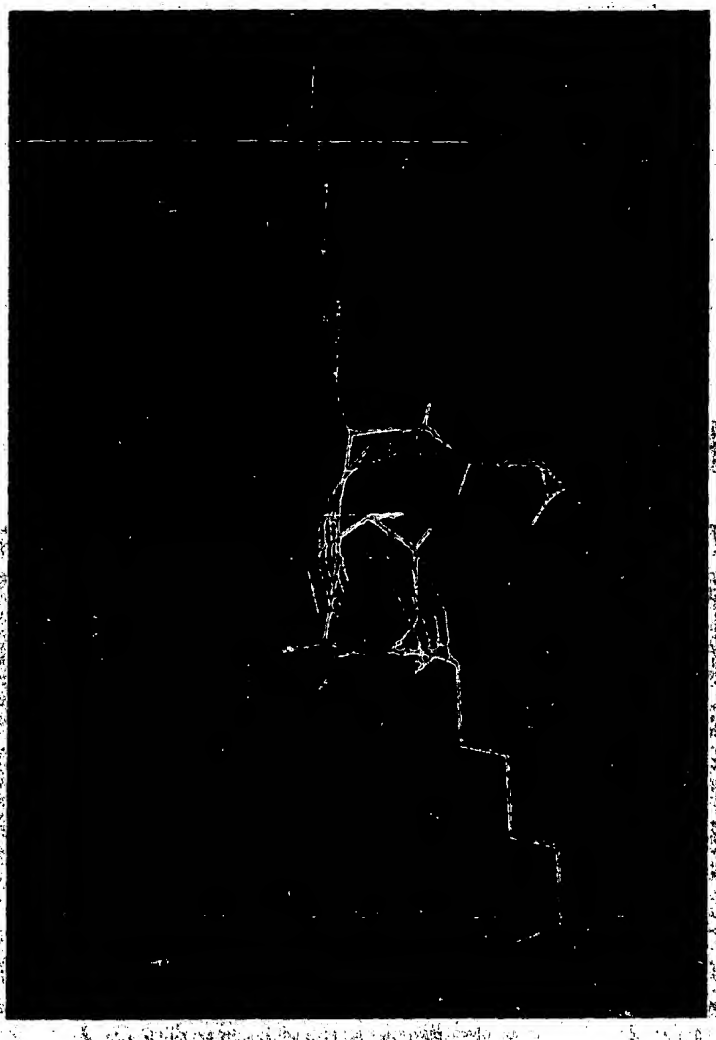

Fig. 2. The Ala-Lys-Thr-Lys-Ala pentapeptide. The threonine residue is mutated to a valine.

the remaining Xs set at 180°; X_1 for threonine was fixed at 288° as before.

Results

The free energy differences for the windows used in the mutation of methanol to ethane are given in Table I. Initial 400 K configurations are carried out for equilibration and 20 batches of 50 K configurations are used to calculate the average free energy change (and standard deviation) for each window. The total F in the forward direction $\lambda = 0 \rightarrow \lambda = 1$ is +8.68 kcal mol^{-1}, i.e. it is unfavourable as a polar hydroxyl group is being replaced by a methyl group. The reverse F is −7.98 kcal/mol. These values assume no hysteresis for the first and last windows. Thus, the average magnitude of 8.3 kcal mol^{-1} should be compared with an experimental value of 6.9 kcal/mol. Moreover, the hysteresis effects are fairly small for windows 2, 3 and 4 as the forward and reverse magnitude of the free energy difference agree within ~0.4 kcal/mol. The first window ($\lambda = 0 \rightarrow \lambda = 0.125$) produces the largest change in free energy, consistent with previous studies (Jorgensen and Ravimohan, 1985).

A similar hydroxyl to methyl mutation is involved in the conversion of a threonine to a valine residue within a tripeptide. A similar number of windows are used but the average free energy change per window is obtained from 25 batches of 50 K configurations. The total forward F is 6.85 kcal mol^{-1}. Again this is positive because of the unfavourable change of an hydroxyl to a methyl group. The total reverse F is −7.20 kcal/mol. The magnitudes of these numbers agree within 0.35 kcal mol^{-1}. The average magnitude is 7.05 kcal/mol, compared with an average

magnitude of 8.33 kcal mol^{-1} found in the ethanol to methane mutation.

In order to test the precision of these estimates, further calculations are undertaken on this threonine to valine mutation within a tripeptide. These tests are only performed on the first windows with $\lambda = 0.125 \rightarrow 0.0$ and $\lambda = 0.125 \rightarrow 0.25$. Using a different starting seed for the random number generator, values of -2.28 ± 0.13 kcal/mol and $+1.56 \pm 0.11$ kcal/mol are obtained for the $\lambda = 0.125 \rightarrow 0.0$ and $\lambda = 0.125 \rightarrow 0.25$ windows respectively. These should be compared with the previous values of -2.56 ± 0.14 and $+1.85 \pm 0.12$ kcal/mol for the same windows. Thus starting with a new seed produces free energy changes which are slightly outside the sum of the errors.

We have also used the N, P, T rather than the N, V, T ensemble. This produces free energy differences of -2.86 ± 0.08 kcal/mol and $+2.31 \pm 0.07$ kcal/mol for windows $\lambda = 0.125 \rightarrow 0.0$ and $\lambda = 0.125 \rightarrow 0.25$ respectively. The precision on these average numbers which are each produced from 28 batches of 50 K configurations is better than that obtained from N, V, T ensembles. Moreover, there are differences of 0.3 and 0.46 kcal mol^{-1} between the same estimates obtained in the N, V, T ensemble. The convergence profiles of the two (N, V, T) simulations and the (N, P, T) simulations for this window are shown in Figure 1.

Due to the fact that the precision of these calculations appears better in the N, P, T ensemble we continued with this ensemble and calculated the free energy difference of a threonine to valine mutation within a pentapeptide but with lysines as neighbouring residues (see Figure 2). We expect the presence of two charged groups to alter considerably the environment of the hydroxyl group. The total forward and reverse changes in free energy are $+6.79$ and -6.43 kcal/mol respectively. The average magnitude of 6.61 kcal/mol is less than that found with the same mutation flanked by neighbouring alanine residues. However, agreement between the magnitude of the forward and backward estimates for each window is disappointing.

Table II. Threonine to valine mutation in Ala-Thr-Ala tripeptide

Coupling parameter		F_{ij}	F_{ji}
λ_i	λ_j		
N, V, T ensemble			
0	0.125	–	−2.561 ± 0.135
0.125	0.25	1.854 ± 0.120	−1.368 ± 0.088
0.25	0.5	1.763 ± 0.163	−1.809 ± 0.237
0.5	0.75	0.359 ± 0.123	−1.122 ± 0.134
0.75	1.0	0.342 ± 0.132	–
N, P, T ensemble			
0	0.125	–	−2.856 ± 0.084
0.125	0.25	2.309 ± 0.067	–
New seed (N, V, T)			
0	0.125	–	−2.277 ± 0.133
0.125	2.5	1.558 ± 0.108	–

The forward (F_{ij}) and reverse (F_{ji}) free energies in kcal/mol. An equilibration for 200 000 Monte Carlo steps is performed initially at constant volume, since undesirable volume expansion can occur to relieve bad contacts. Data are then collected at constant pressure for the next 1.3×10^6 Monte Carlo steps.

Table III. Threonine to valine mutation in Ala-Lys-Thr-Lys-Ala pentapeptide

Coupling parameter			
λ_i	λ_j	F_{ij}	F_{ji}
0	0.125	+2.345 ± 0.089	−2.614 ± 0.133
0.125	0.25	1.835 ± 0.120	−1.057 ± 0.078
0.25	0.5	1.337 ± 0.129	−2.400 ± 0.107
0.5	0.75	1.465 ± 0.090	−0.219 ± 0.092
0.75	1.0	0.159 ± 0.090	−0.064 ± 0.095
Total		6.823	−6.354

The forward (F_{ij}) and reverse (F_{ji}) free energies in kcal/mol. An initial equilibrium was performed for 100 000 Monte Carlo steps at constant volume to ensure no undesirable expansion occurs to relieve bad contacts. Data were collected at constant pressure for the next 1.1×10^6 steps.

Discussion

Pearlman and Kollman (1989a) comment that one of the main advantages of free energy simulations is that they yield a single number which can be compared to experiment. However, they also point out that this can be the main shortcoming of the method as it is often difficult to know how reliable this one number is. Sometimes repeating the simulation with small and reasonable changes in procedure (e.g. a new starting seed in a Monte Carlo run or a small change in a dynamics trajectory) can lead to appreciably different answers. This is particularly true for larger systems, where there will be a more complicated energy surface, and more local minima. The fact that not all regions of phase space are accessible, which results in incomplete sampling, remains a problem with free energy simulations. Adaptive importance sampling (Mezei, 1987) may be of use in this regard.

We have carried out a free energy 'window' perturbation simulation for the mutation of an OH group to a CH_3 group in three different local environments. The same window intervals are used in all cases. While the agreement between forward and backward run is satisfactory for the methanol/ethane mutation, there are large hysteresis effects for window 4 (see Table II) in the Ala-Thr-Ala tripeptide. For the pentapeptide there are significant hysteresis problems for windows 2, 3 and 4. In this case, the hysteresis for window 1, which accounts for the largest free energy change, is quite small, although outside the error bars (+2.35 ± 0.09 and 2.61 ± 0.13 kcal/mol). Somewhat surprisingly, the total free energy changes for the forward and backward simulations are in reasonable agreement. This effect has been observed by Pettitt (1989). Using slow growth methods, Pettitt noticed that sometimes the total final answers are in better accord than answers at any time in the run other than at the end points, although in principle any particular time could have been the physical end point. Perhaps a good measure of overall error should be the largest discrepancy between F_{ij} and F_{ji} at any window.

For the tripeptide, comparison of an N, V, T simulation with two different starting seeds and an N, P, T simulation show reasonable agreement (Figure 1), although the averages are just outside the error bars. Figure 1 suggests that the agreement is improving with the length of the simulation.

Despite their tremendous potential for biomolecular simulation free energy perturbation calculations do suffer from the so called multiple minimum problem. Detailed studies on simple systems reveal how robust the method is and allows problems to be highlighted, so the techniques can be used with greater confidence on larger biological systems. Our results suggest that the magnitude of the free energy of mutating an OH group to a CH_3

group in aqueous solution is dependent on the local environment, as it becomes less, on going from the methanol/ethane system to the Ala-Thr-Ala tripeptide and finally the Ala-Lys-Thr-Lys-Ala pentapeptide. The accuracy of the results for the more complicated systems decreases. A different choice of window intervals may lead to better sampling for the larger systems and it would be interesting to use the recent method of dynamically modified windows (Pearlman and Kollman, 1986b) for these systems. Alternatively longer simulations for the windows showing large hysteresis effects may be necessary.

Comparison between our free energy simulations and experimental data is important. For this reason, we initially chose to look at the ethane to methanol transition for which experimental data exist. There appears to be little experimental data on changes in proteins with which we can make a direct comparison. The work of Fersht and co-workers (Kellis *et al.*, 1988; Serrano and Fersht, 1989) and Alber *et al.* (1987) on the thermodynamic stability of mutations to the proteins barnase and phage T4 lysozyme respectively cannot be immediately transferred to the systems we have studied, mainly but not entirely because these experimental results refer to changes essentially within the protein and not in regions with large accessibility to solvent.

In summary, this study highlights the difference in free energy between solvent–apolar and solvent–polar group interactions. The deletion of an apolar group and simultaneous creation of a polar atom (e.g. valine to threonine transition) in aqueous solution appears to be dependent on the local environment and enthalpically driven. Comparison between the total free energy change and the changes to the enthalpic terms shows that there is considerable restriction of water molecules (i.e. unfavourable entropic contributions) around apolar groups as well as around polar groups. This latter effect is presumably due to translational and orientational loss of freedom on formation of solvent–solute hydrogen bonds.

Acknowledgements

We would like to thank the Science and Engineering Research Council for support under project grant GR/D/86683 + GR/E/47025.

References

Alber,T., Dao-Pin,S., Wilson,K., Wozniak,J.A., Cook,S.P. and Matthews,B.W. (1987) *Nature*, **330**, 41–46.
Bash,P.A., Singh,U.C., Langridge,R. and Kollman,P.A. (1987a) *Science*, **236**, 564–568.
Bash,P.A., Singh,U.C., Brown,F.K., Langridge,R. and Kollman,P.A. (1987b) *Science*, **23**, 574–575.
Ben-Naim,A. and Marcus,Y. (1984) *J. Chem. Phys.*, **81**, 2106.
Beveridge,D.L. and DiCapua,F.M. (1989) *Annu. Rev. Biophys. Biophys. Chem.*, in press.
Brooks,C.L.,III (1989) In van Gursteren,W. and Weiner,P. (eds), *Computation of Free Energy for Biomolecular Systems*. Escom Science Publishers, Leiden, pp. 73–88.
Jorgensen,W.L. (1985) *J. Chem. Phys.*, **79**, 926.
Jorgensen,W.L. (1989) *Acc. Chem. Res.*, **22**, 184–189.
Jorgensen,W.L. and Ravimohan,C. (1985) *J. Chem. Phys.* **83**, 3050–3054.
Jorgensen,W.L. and Tirado Rives,J. (1988) *J. Am. Chem. Soc.*, **110**, 1657.
Kellis,J.T., Nyberg,K., Sali,D. and Fersht,A.R. (1988) *Nature*, **333**, 784–786.
Lybrand,T.P., Ghosh,I. and McCammon,J.A. (1985) *J. Am. Chem. Soc.*, **107**, 7793–7794.
McGregor,M., Islam,S.A. and Sternberg,M.J.E. (1987) *J. Mol. Biol.*, **198**, 295.
Mezei,M. (1987) *J. Comp. Chem.*, **68**, 237.
Mezei,M. and Beveridge,D.L. (1986) *Ann. NY Acad. Sci.*, **482**, 1–23.
Pearlman,D.A. and Kollman,P.A. (1989a) In van Gursteren,W. and Weiner,P. (eds), *Computation of Free Energy for Biomolecular Systems*. Escom Science Publishers, Leiden, pp. 101–119.
Pearlman,D.A. and Kollman,P.A. (1989b) *J. Chem. Phys.*, **90**, 2460–2470.
Pettit,B.M. (1989) In van Gursteren,W. and Weiner,P. (eds), *Computation of Free Energy for Biomolecular Systems*. Escom Science Publishers, Leiden, pp. 94–100.
Rao,S.N., Singh,U.C., Bash,P.A. and Kollman,P.A. (1987) *Nature*, **328**, 551–553.
Serrano,L. and Fersht,A.R. (1989) *Nature*, **342**, 296–299.
Singh,U.C., Brown,F.K., Bash,P.A. and Kollman,P.A. (1987) *J. Am. Chem. Soc.*, **109**, 1607–1614.
Thomas,G., Russell,A.J. and Fersht,A.R. (1985) *Nature*, **318**, 375–376.
Wolfenden,R., Anderson,L., Cullis,P.M. and Southgate,C.C.B. (1981) *Biochemistry*, **20**, 849–855.
Zamytnin,A.A. (1984) *Annu. Rev. Biophys. Bioengng.*, **13**, 145–165.

Received on October 4, 1989; accepted on January 15, 1990